

### **Abstract of the Disclosure**

The present invention relates to systems, methods, and computer program products for the analysis of gene expression data, especially data that have been acquired using microarray technologies. The present invention relates to methods for partitioning a set of genes into clusters, based on the similarity of the genes' rates of messenger RNA synthesis. The present invention also relates to methods for annotating clusters with words or phrases that are extracted from documents associated with genes in the clusters. The present invention also relates to methods for evaluating the quality of clustering based on the extent to which documents associated with genes in a cluster collectively distinguish that cluster from all the other clusters, as well as the extent to which some words and phrases, present in documents associated with genes in the cluster, collectively distinguish that cluster from all the other clusters.

Patent 6,838,438